ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 6 — Information Theory and Coding

Solutions 3 — Oct. 11, 2011

PROBLEM 1.

(a) Let X be the number of tosses until the first head appears in a sequences of independent coin tosses, suppose the coin lands heads with probability p, and tails with probability q. Then $X = n$ if and only if the first $n-1$ tosses are tails and the last one is a head. Thus $\Pr(X = n) = pq^{n-1}$, $n = 1, 2, \ldots$. Then

$$\begin{aligned} H(X) &= -\sum_{n=1}^{\infty} pq^{n-1} \log(pq^{n-1}) \\ &= -\left[\sum_{n=0}^{\infty} pq^n \log p + \sum_{n=0}^{\infty} npq^n \log q\right] \\ &= \frac{-p \log p}{1-q} - \frac{pq \log q}{p^2} \\ &= \frac{-p \log p - q \log q}{p} \\ &= H(p)/p \text{ bits.} \end{aligned}$$

If $p = 1/2$, then $H(X) = 2$ bits.

(b) One possible questioning strategy is to ask the questions 'Is $X = 1$?', 'Is $X = 2$?', 'Is $X = 3$?', ..., stopping whenever a 'yes' answer is given. The number of questions asked when $X = n$ is exactly n, and thus the expected number of questions asked is $\sum_{n=1}^{\infty} n(1/2^n) = 2$.

Since this equals $H(X)$ this strategy cannot be improved upon.

PROBLEM 2. For any RV X and any function f, $H(X) \geq H(Y)$ where $Y = f(X)$. [Proof: Observe that $H(Y|X) = 0$, and write $H(X, Y) = H(X) + H(Y|X) = H(X)$, and $H(X, Y) = H(Y) + H(X|Y) \geq H(Y)$.] With this in mind, since the run lengths are a function of $X_1, \ldots, X_n$, $H(\mathbf{R}) \leq H(X_1, \ldots, X_n)$. Any X_i together with the run lengths determine the entire sequence $X_1, \ldots, X_n$. Hence

$$\begin{aligned} H(X_1, \ldots, X_n) &= H(X_i, \mathbf{R}) \\ &= H(\mathbf{R}) + H(X_i|\mathbf{R}) \\ &\leq H(\mathbf{R}) + H(X_i) \\ &\leq H(\mathbf{R}) + 1. \end{aligned}$$

PROBLEM 3.

(a) $H(X) = \frac{2}{3} \log \frac{3}{2} + \frac{1}{3} \log 3 = 0.918 \text{ bits} = H(Y)$.

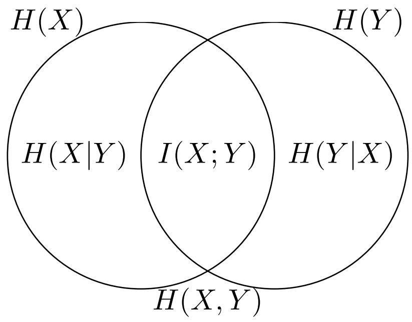
(b) $H(X|Y) = \frac{1}{3} H(X|Y = 0) + \frac{2}{3} H(X|Y = 1) = 0.667 \text{ bits} = H(Y|X)$.

(c) $H(X, Y) = 3 \times \frac{1}{3} \log 3 = 1.585$ bits.

(d) $H(Y) - H(Y|X) = 0.251$ bits.

(d) $I(X;Y) = H(Y) - H(Y|X) = 0.251$ bits.

(f)



PROBLEM 4.

(a) Using the chain rule for mutual information,

$$I(X,Y;Z) = I(X;Z) + I(Y;Z \mid X) \geq I(X;Z),$$

with equality iff $I(Y;Z \mid X) = 0$, that is, when Y and Z are conditionally independent given X.

(b) Using the chain rule for conditional entropy,

$$H(X,Y \mid Z) = H(X \mid Z) + H(Y \mid X,Z) \geq H(X \mid Z),$$

with equality iff $H(Y \mid X,Z) = 0$, that is, when Y is a function of X and Z.

(c) Using the chain rule for mutual information,

$$I(X;Z \mid Y) + I(Z;Y) = I(X,Y;Z) = I(Z;Y \mid X) + I(X;Z),$$

and therefore

$$I(X;Z \mid Y) = I(Z;Y \mid X) - I(Z;Y) + I(X;Z).$$

We see that this inequality is actually an equality in all cases.

(d) Using first the chain rule for entropy and then the definition of conditional mutual information,

$$\begin{aligned} H(X,Y,Z) - H(X,Y) &= H(Z \mid X,Y) = H(Z \mid X) - I(Y;Z \mid X) \\ &\leq H(Z \mid X) = H(X,Z) - H(X), \end{aligned}$$

with equality iff $I(Y;Z \mid X) = 0$, that is, when Y and Z are conditionally independent given X.

Problem 5.

(a) Since the lengths prescribed satisfy the Kraft inequality, an instantaneous code can be used for the final stage of encoding the intermediate digits to binary codewords. In this case, each stage of the encoding is uniquely decodable, and thus the overall code is uniquely decodable.

(b) The indicated source sequences have probabilities $0.1, (0.9)(0.1), (0.9)^2(0.1), (0.9)^3(0.1), \ldots, (0.9)^7(0.1), (0.9)^8$. Thus,

$$\bar{N} = \sum_{i=1}^{8} i(0.1)(0.9)^{i-1} + 8(0.9)^8 = 5.6953.$$

(c)

$$\bar{M} = 1(0.9)^8 + 4[1 - (0.9)^8] = 2.7086.$$

(d) Let $N(i)$ be the number of source digits giving rise to the first i intermediate digits. For any $\epsilon > 0$

$$\lim_{i\to\infty} \Pr\left[\left|\frac{N(i)}{i} - \bar{N}\right| > \epsilon\right] = 0.$$

Similarly, let $M(i)$ be the number of encoded bits corresponding the the first i intermediate digits. Then

$$\lim_{i\to\infty} \Pr\left[\left|\frac{M(i)}{i} - \bar{M}\right| > \epsilon\right] = 0.$$

From this, we see that for any $\epsilon > 0$,

$$\lim_{i\to\infty} \Pr\left[\left|\frac{M(i)}{N(i)} - \frac{\bar{M}}{\bar{N}}\right| > \epsilon\right] = 0,$$

and that for a long source sequence the number of encoded bits per source digit will be $\bar{M}/\bar{N} = 0.4756$.

The average length of the Huffman code encoding 4 source digits at a time is 1.9702, yielding $1.9702/4 = 0.49255$ encoded bits per source digit.

For those of you puzzled by the fact that the 'optimum' Huffman code gives a worse result for this source than the run-length coding technique, observe that the Huffman code is the optimal solution to a mathematical problem with a given message set, but the choice of a message set can be more important than the choice of code words for a given message set.